

共分散構造分析の数理

狩野 裕
大阪大学 基礎工学研究科



1

2

内容

第一部 基礎編

1. 次元縮小と誤差分離
2. 尺度とSEM
3. 離散変数のモデル
 - 3.1 離散従属観測変数
 - 3.2 離散独立観測変数
 - 3.3 離散潜在変数

第二部 数理編

1. 共分散構造
RAM
具体例
2. 適合度指標
3. 共分散構造分析における統計的推測



3

モデリングの時代

- 相関(構造)に主な興味
 - 構造方程式モデリング
 - CALIS
- 平均(構造)に主な興味
 - 混合モデル
 - MIXED



第一部 基礎編

1. 次元縮小と誤差分離
2. 尺度とSEM
3. 離散変数のモデル



4

5

SEMの特徴

- 潜在変数の導入
 - 次元縮小
 - 誤差を分離して正確な分析を提供
- 因果分析
 - パス解析
- 多様なモデル表現1
 - 因子モデル+パス解析モデル
 - 多母集団同時分析
 - 平均構造分析
- 多様なモデル表現2
 - PARAFAC
 - 直積モデル
- 多様なモデル表現3
 - 二値データの因子分析, IRT
 - ロジスティック回帰分析
- 多様なモデル表現4
 - 潜在曲線モデル
 - 多段抽出モデル
 - 有限混合モデル



1. 次元縮小と誤差分離



6

7

タレント好感度データ

- 狩野(1998) 豊田秀樹(編著)共分散構造分析 [事例編]. 北大路書房.
- 一週間おいて2回調査. 反復測定
- N=86, 女子学生
- 目的
 - タレント好感度の構造を探る

好きなタレント大調査98+

学部 _____ 学年 _____ 男女 _____ 自分だけ分かる記号 _____

○ 当てはまる項目に丸をつけて下さい ○

山口智子	大好き	好き	どちらかといえば好き	どちらでもない	どちらかといえば嫌い	嫌い	大嫌い
広末涼子	大好き	好き	どちらかといえば好き	どちらでもない	どちらかといえば嫌い	嫌い	大嫌い
宮澤あま	大好き	好き	どちらかといえば好き	どちらでもない	どちらかといえば嫌い	嫌い	大嫌い

8

データの要約

観測変数の共分散(下三角), 相関(上三角), 平均

	1松	2松	1松本	2松本	1森口	2森口
1松たか子	2.10	0.89	0.18	0.21	0.21	0.23
2松たか子	1.78	1.91	0.17	0.23	0.25	0.21
1松本明子	0.37	0.32	1.95	0.79	0.49	0.44
2松本明子	0.39	0.40	1.38	1.59	0.50	0.41
1森口博子	0.37	0.41	0.82	0.75	1.44	0.81
2森口博子	0.36	0.33	0.67	0.57	1.07	1.22
平均	3.83	3.79	3.80	3.94	4.36	4.22

- 平均
 - 松と松本が近い
- 相関
 - 松本と森口が近い

9

誤差分離と次元縮小

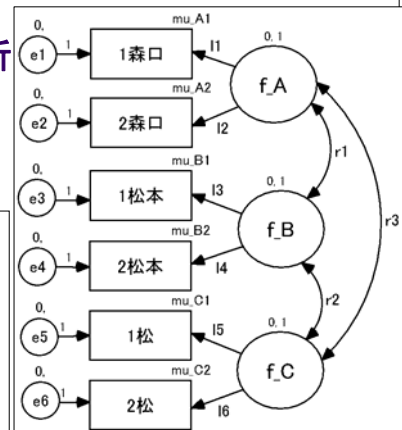
- 回答の揺れ
 - 真の好感度Fに誤差eが伴って回答
 - 誤差を分離しバイアスのない推測
- 次元縮小
 - 真の好感度は一つ
 - 各回の回答はF+e1, F+e2として観測
 - 誤差を取り除くことで次元縮小
- 真の好感度F(潜在変数)の導入で誤差分離と次元縮小を図る

10

検証的因子分析モデル (CFA)

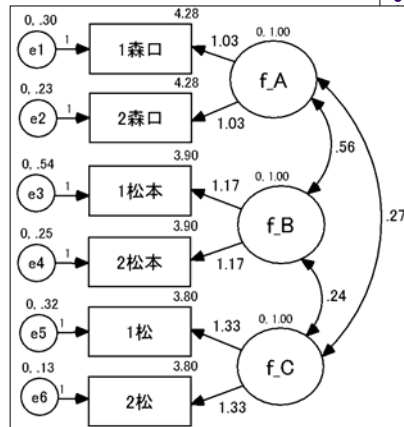
等値制約
 $I1=I2$
 $I3=I4$
 $I5=I6$
 $\mu_{A1}=\mu_{A2}$
 $\mu_{B1}=\mu_{B2}$
 $\mu_{C1}=\mu_{C2}$

(自由に)等値制約を課することができる
 →柔軟なモデリング



11

推定結果

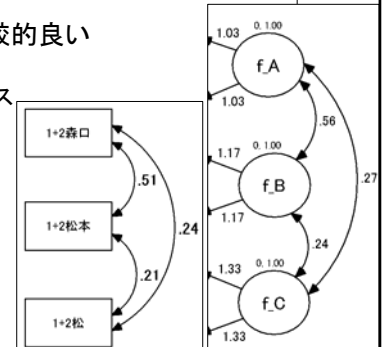


タレント好感度の分析
 カイ2乗値=13.952
 df=12, p値=.304
 GFI=.967, CFI=.994
 RMSEA=.044

12

合計得点法

- 合計得点法は比較的良いパフォーマンス
 - 1割程度のバイアス
- 信頼性が高い
 - 項目間相関
 - 0.8~0.9
 - α 係数
 - 森口: 0.89
 - 松本: 0.88
 - 松: 0.94



信頼性による修正

- 希薄化された相関係数を信頼性によって修正できる
 - 合計得点
 - 単純相関

$$\begin{aligned} \text{Cor}(\text{森口, 松本}) &= \frac{0.51}{\sqrt{0.89 \times 0.88}} \approx 0.56 \\ \text{Cor}(\text{松本, 松}) &= \frac{0.21}{\sqrt{0.88 \times 0.94}} \approx 0.24 \\ \text{Cor}(\text{松, 森口}) &= \frac{0.24}{\sqrt{0.94 \times 0.89}} \approx 0.37 \end{aligned}$$



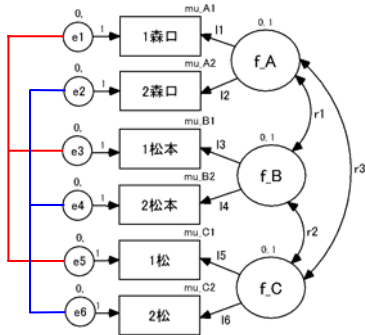
本分析についての2つの疑問

- **2指標で十分なのか？**
 - 十分！
 - 潜在変数が真の好感度を表すことが明確
 - 誤差分離が目的ならば2指標で問題なし
- **信頼性による修正はあまり行われたい？**
 - 理論は散見
 - 実証研究では見かけない
 - 理由
 - α係数が不正確
 - モデストな分析
 - モデル適合の必要性 → 次スライド



モデル適合の必要性

- 信頼性による修正
 - 想定外の相関があると不当に過大推定される
 - 誤差相関
 - e.g. 方法因子
 - 誤差相関のないCFAモデル適合の吟味が必要
 - SEMIによる分析
 - 本データでは誤差相関なし



2. 尺度とSEM

尺度開発におけるSEMの役割



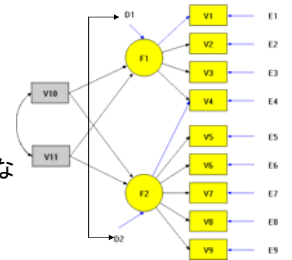
EFAとCFA

- EFA
 - 尺度開発における主たる道具
 - 大規模な探索的項目選択を伴う尺度開発にはCFAは向かない
 - e.g., 100項目→40項目
- CFA
 - 尺度の精緻化
 - 尺度の再吟味
 - 尺度 (構成概念)間の関係の分析
 - 相関分析, 回帰分析

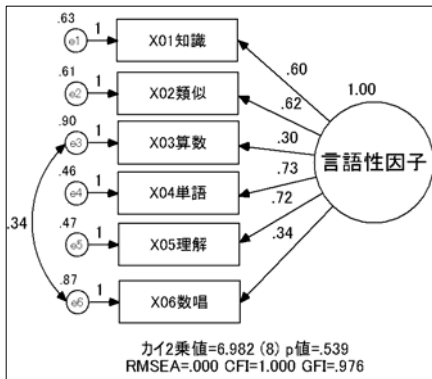


CFAによる尺度吟味

- 適合度の吟味
 - 安易に尺度項目を落とさない
- クロスのパスの有意性
 - 想定外の因子負荷
- 誤差共分散
 - 属性変数の影響を取り除く
 - 通例は, 属性変数の影響がないことを仮定



WISC-Rの分析



PCA, EFAのパフォーマンス

言語性因子の因子負荷量			
	誤差共分散あり	誤差共分散なし	PCA
X01知識	0.60	0.61	0.70
X02類似	0.62	0.63	0.72
X03算数	0.30	0.35	0.52
X04単語	0.73	0.71	0.73
X05理解	0.72	0.70	0.73
X06数唱	0.34	0.39	0.55
カイ2乗値	6.982(8)	22.111(9)	
p値	0.539	0.009	
GFI	0.976	0.931	0.739
rho'	0.703	0.746	0.823
α	0.747	0.747	0.747

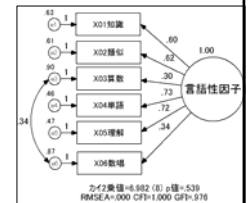
信頼性係数

- 尺度と尺度項目
 $X = X_1 + \dots + X_p$
- 1因子分析モデルにもとづく信頼性
 $X_i = \lambda_i f + e_i \quad (i=1, \dots, p)$
 $X = \left(\sum_{i=1}^p \lambda_i \right) f + \sum_{i=1}^p e_i$
$$\rho = \frac{\left(\sum_{i=1}^p \lambda_i \right)^2}{\left(\sum_{i=1}^p \lambda_i \right)^2 + \sum_{i=1}^p V(e_i)}$$
- 誤差共分散がある場合
$$\rho' = \sigma = \frac{\left(\sum_{i=1}^p \lambda_i \right)^2}{\left(\sum_{i=1}^p \lambda_i \right)^2 + \sum_{i=1}^p V(e_i) + \sum_{i \neq j} Cov(e_i, e_j)}$$
- Cronbachのα係数
$$\alpha = \frac{p}{p-1} \left[1 - \frac{\sum_{i=1}^p V(X_i)}{V(X)} \right]$$

$$= \frac{p}{p-1} \left[\frac{\sum_{i \neq j} Cov(X_i, X_j)}{V(X)} \right]$$
- 1因子モデルが正しく且つ $\lambda_1 = \dots = \lambda_p$ のとき
 $\alpha = \rho = \rho'$
- 1因子モデルが正しいとき
 $\alpha \leq \rho = \rho'$
- $Cov(e_i, e_j) \neq 0$ のとき
 $\alpha \neq \rho \neq \rho'$

測定モデルにおける指標の数

- 誤差分離の場合
 - 因子の相関分析・回帰分析
 - 2でよい
 - 合計(尺度)得点の利用
 - 十分な信頼性を確保
- 構成概念を規定する場合
 - 因子の相関分析・回帰分析
 - 共通因子の解釈に多義性が残らないように
 - 合計(尺度)得点の利用
 - 内容的妥当性が確保されるように
 - 十分な信頼性を確保



3. 離散変数のモデル

23

3.1 離散従属観測変数

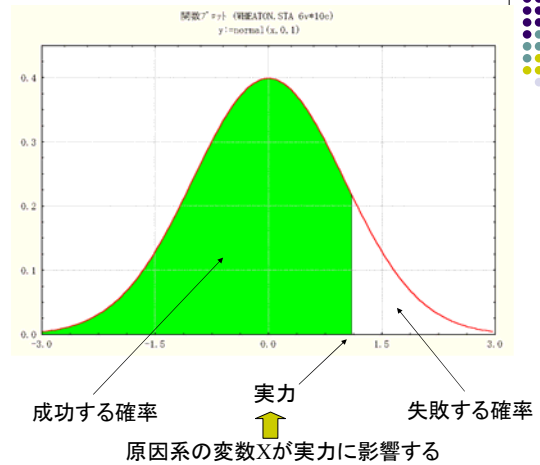
- 例... 二値データの因子分析
- 5件法(以上)は正規分布とみなすことが多い
 - 変数のヒストグラムの吟味が必須
- 3件法と4件法はグレーゾーン
 - 多値ロジスティック回帰モデル
 - プログラムはMplusがよい
- 2件法は次スライド

25

2件法の場合

- 2件法(二値変数)に対しては特殊な方法が必要
 - 二値変数を連続変数で予測するのはまずい
 - $y = \beta_0 + \beta_1 x + \varepsilon$
 - $Y \sim 0.5, -2.1, 1.5$ などに意味がない
 - 生起確率 $P(y=1)$ が独立変数に依存すると考える
 - ロジスティック回帰モデルと同じ考え方
 - $\text{logit } P(y=1|x) = \beta_0 + \beta_1 x$
 - $\Phi^{-1}(P(y=1|x)) = \beta_0 + \beta_1 x$
- プログラムはMplusがよい

26



27

3.2 離散独立観測変数

- 回帰分析と同じ考え方
 - 独立変数を与えた下で従属変数の条件付分布を考える

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

$$\begin{cases} x_1 = 0 & \text{for male} \\ x_1 = 1 & \text{for female} \end{cases}$$

x_2 : 連続変数

β_1 は平均の性差を表す

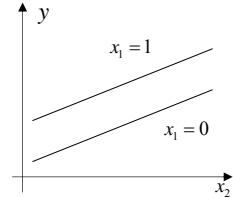
$$\begin{aligned} E[y | x_1 = 1, x_2] - E[y | x_1 = 0, x_2] \\ = (\beta_0 + \beta_1 \cdot 1 + \beta_2 x_2) - (\beta_0 + \beta_1 \cdot 0 + \beta_2 x_2) \\ = \beta_1 \end{aligned}$$

独立観測変数は条件を付けて考える

$$y |_{x_1, x_2} \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2, V(e))$$

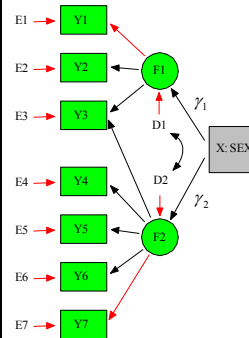
仮定

- x_1 と x_2 の交互作用はない
- 男女で誤差分散に違いはない



28

SEMの場合



構造方程式

$$y = \mu + \Lambda f + e$$

$$f = \gamma X + d$$

\Rightarrow

$$y = \mu + \Lambda \gamma X + \Lambda d + e$$

条件付分布

$$y |_{X=x} \sim N(\mu + \Lambda \gamma x, V(\Lambda d + e))$$

$$E(f | X = x) = \gamma x$$

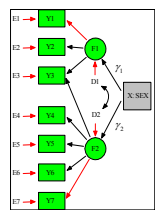
$$V(f | X = x) = V(d)$$

$$V(\Lambda d + e) = \Lambda V(d) \Lambda' + V(e)$$

29

考察 1/2

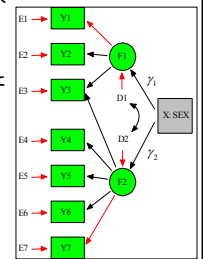
- Xからのパス(係数)は因子平均を表す
 - $X=0$ と $X=1$ の集団の平均差
- $V(d)$
 - (X の影響を取り除いた後の) 因子の共分散行列
- $V(y|X=x) = V(\Lambda d + e)$
 - (X の影響を取り除いた後の) y の共分散行列
 - $X=0$ と $X=1$ の集団によらず一定



30

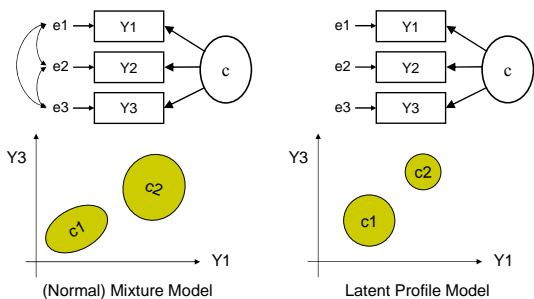
考察 2/2

- 平均構造
 - $E(y|X=x) = \mu + \Lambda \gamma x$
 - $X=0$ と $X=1$ の集団によらず y の切片は共通
 - 集団による因子平均の差が y の平均に反映
- 多母集団同時分析と同等
 - y の切片と共分散構造を完全に等値
 - 回帰分析や(多変量)分散分析でも課された仮定
 - 共分散行列の等質性
 - 暗黙の了解? 罪は重くない?
 - 因子平均が異なるモデル
- 注意: 適合度は異なる



31

3.3 離散潜在変数



32

離散潜在変数とは

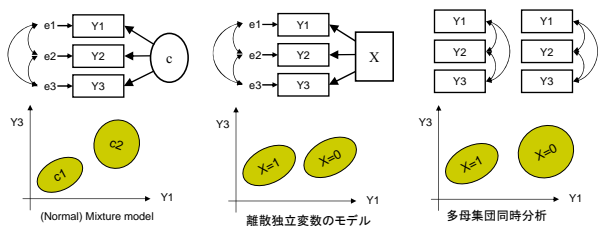
- 離散潜在変数は潜在クラスを表す
- 潜在クラス
 - 平均や共分散行列が異なるいくつかの部分母集団
 - どちらか同じであってもよい
 - 事前に部分母集団は特定できない
 - 「潜在」の意味するところ
 - クラスタ分析と類似、教師なしの分類、探索的分類
- 得られたデータは
 - 異なる特性をもつ潜在クラスを合併したもの
 - 統計学では(有限)混合モデルという
 - SEMのcontextでは
 - Latent Variable Mixture Modeling
 - Structural equation mixture model
- プログラム
 - M-plus by Muthen & Muthen

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \pi_1 N_p(\mu_1, \Sigma_1) + (1 - \pi_1) N_p(\mu_2, \Sigma_2)$$

33

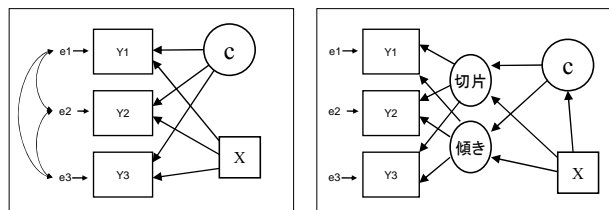
多母集団同時分析や離散独立変数モデルとの違い

- 部分母集団を規定する要因が未知
 - 男女, 国籍
- モデル規定にも多少の違いあり

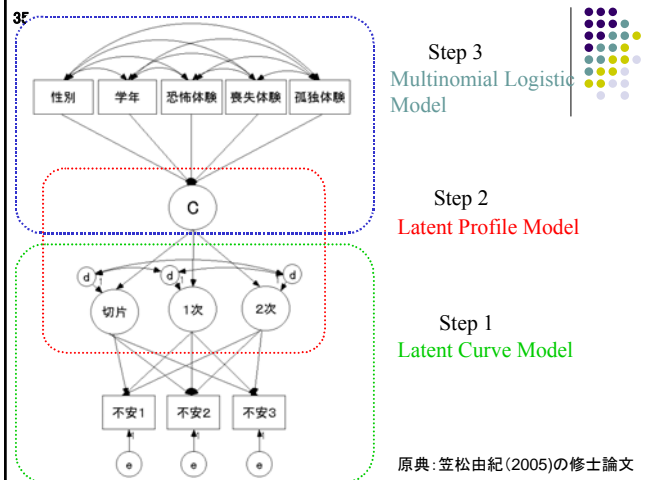


34

こんなことも



35



36

モデル式

Latent Curve Model + Latent Profile Model
 Latent Class: $c = 1, \dots, K (= 24)$

$$\begin{bmatrix} X_1^{(c)} \\ X_2^{(c)} \\ X_3^{(c)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} f_{int}^{(c)} \\ f_{lin}^{(c)} \\ f_{qua}^{(c)} \end{bmatrix} + e$$

$$E \begin{bmatrix} f_{int}^{(c)} \\ f_{lin}^{(c)} \\ f_{qua}^{(c)} \end{bmatrix} = \begin{bmatrix} \eta_{int}^{(c)} \\ \eta_{lin}^{(c)} \\ \eta_{qua}^{(c)} \end{bmatrix}$$

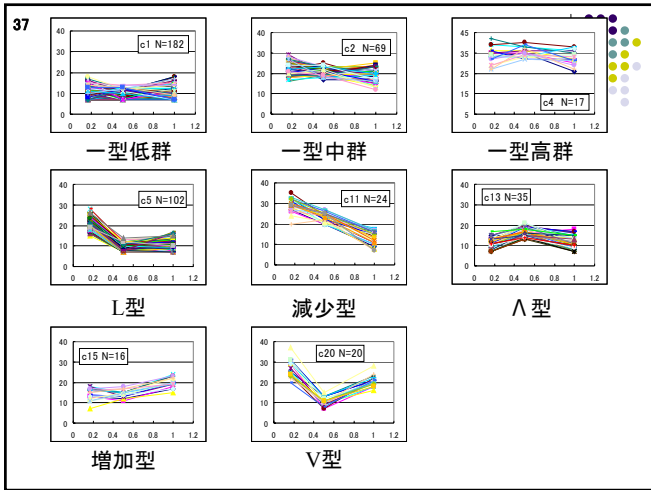
$$V \begin{bmatrix} f_{int}^{(c)} \\ f_{lin}^{(c)} \\ f_{qua}^{(c)} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}$$

Multinomial Logistic Model

$$\log \frac{P(C=c|\mathbf{x})}{P(C=K|\mathbf{x})} = \alpha^{(c)} + (\boldsymbol{\gamma}^{(c)})^T \mathbf{x} \quad (c=1, \dots, K)$$

or equivalently

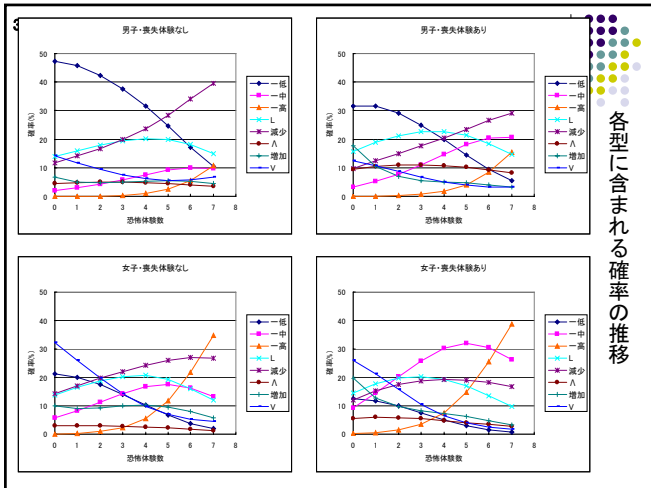
$$P(C=c|\mathbf{x}) = \frac{e^{\alpha^{(c)} + (\boldsymbol{\gamma}^{(c)})^T \mathbf{x}}}{\sum_{k=1}^K e^{\alpha^{(k)} + (\boldsymbol{\gamma}^{(k)})^T \mathbf{x}}} \quad (c=1, \dots, K)$$



36

男子 喪失 恐怖	一型			L型			減少型		Λ型	増加型		V型		N
	低	中	高	小低	小高	大	小	大		小	大	小	大	
0	47.4			12.5			7.8						12.0	6
1	45.7			14.0			9.5						9.6	31
2	42.3			15.1			11.2	5.6					7.4	99
3	37.6	5.7		15.6			12.5	7.3					5.5	79
4	31.7	7.4		15.3			13.3	10.2						33
5	24.7	9.1		13.8			13.1	15.2						8
6	17.2	10.1	5.5	11.2			11.6	22.4						2
7	10.5	9.8	10.8				7.9	30.8						1
0	31.6			13.9			5.8		9.3		16.3		9.5	0
1	31.0	5.3		16.2			7.4	5.1	10.5		8.4		7.9	6
2	29.1	7.8		17.3			8.6	6.5	11.0				6.0	48
3	25.0	11.0		17.2			9.3	8.3	11.1					48
4	19.9	14.7		16.0			9.3	11.0	10.7					40
5	14.5	18.1		13.5			8.6	14.8	10.2					7
6	9.4	20.4	8.4	10.2			7.1	19.5	9.3					1
7	5.4	20.5	15.5				5.1	24.1	8.2					0
N	141	32	3	63	7	8	41	30	28	11	8	15	9	406
女子 喪失 恐怖	一型			L型			減少型		Λ型	増加型		V型		N
低	中	高	小低	小高	大	小	大	小		大	小	大		
0	21.3	5.7		10.5			9.5	5.7			5.7		25.8	6.3
1	19.9	9.3		11.5			11.3	5.8			5.5		20.2	5.8
2	17.4	11.3		11.6	6.0		12.4	7.2			6.9		14.7	5.0
3	14.1	14.3		10.9	7.9		12.7	9.4			8.0		9.9	75
4	10.3	16.7	5.6	9.3	9.5		11.7	12.4			8.5		6.0	34
5	6.8	17.5	11.8	7.1	10.2		9.7	16.3			8.2			15
6	16.3	21.9		9.5			7.0	20.1			6.8			1
7	13.2	34.7					7.6	22.5						0
0	12.8	9.4		10.4			6.3	5.8	5.5		16.5		16.1	7.8
1	11.9	14.5		11.4	5.3		7.5	7.7	6.0		8.5		14.2	7.1
2	9.9	20.3		11.0	7.3		7.9	9.6	5.9	5.2			9.5	36
3	7.4	25.9		9.8	9.0		7.5	11.3	5.4	5.7			6.2	48
4		30.2	7.6	7.5	9.9		6.3	13.0		5.6				30
5		32.0	14.7	5.2	9.7		5.2	14.3						15
6		30.5	25.4		8.5			15.1						3
7		28.1	38.7		6.5			14.9						1
N	41	68	14	39	30	4	40	32	12	23	7	28	15	353

説明変数を与えたもので、各群に含まれる確率



第一部終了

Any questions are welcome!

40

第二部 数理編

1. 共分散構造の導出
2. 適合度指標
3. 共分散構造分析における統計的推測

41

1. 共分散構造の導出

42

43

共分散構造とは

- (推定すべき)母数 (parameter)
 - θ で表す(ベクトル)
 - パス係数(偏回帰係数)と独立変数の分散・共分散
 - (独立変数の平均, 従属変数の切片)
- 共分散構造 (covariance structure)
 - 観測変数の共分散行列を母数の関数で表したものを $V(X)=\Sigma(\theta)$
- 共分散構造を導出する手順が必要
 - モデルが多様
 - どんなモデルに対しても適用可能
 - RAM (Reticular Action Model) が分かりやすい



44

共分散構造の導出:一般公式

$$\eta = B\eta + \Gamma\xi + e \Leftrightarrow \begin{bmatrix} \eta \\ \xi \end{bmatrix} = \begin{bmatrix} B & \Gamma \\ O & O \end{bmatrix} \begin{bmatrix} \eta \\ \xi \end{bmatrix} + \begin{bmatrix} e \\ \xi \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} \eta \\ \xi \end{bmatrix} = \begin{bmatrix} I-B & -\Gamma \\ O & I \end{bmatrix}^{-1} \begin{bmatrix} e \\ \xi \end{bmatrix} \Leftrightarrow x = G \begin{bmatrix} \eta \\ \xi \end{bmatrix} = G \begin{bmatrix} I-B & -\Gamma \\ O & I \end{bmatrix}^{-1} \begin{bmatrix} e \\ \xi \end{bmatrix}$$

$$\Rightarrow V(x) = G \begin{bmatrix} I-B & -\Gamma \\ O & I \end{bmatrix}^{-1} \begin{bmatrix} V(e) & O \\ O & V(\xi) \end{bmatrix} \left(\begin{bmatrix} I-B & -\Gamma \\ O & I \end{bmatrix}^{-1} \right)^T G^T = \Sigma(\theta)$$



45 パス解析モデル

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & b_{13} \\ 0 & 0 & b_{23} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} 0 \\ b_{24} \\ b_{34} \end{bmatrix} X_4 + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & b_{13} & 0 \\ 0 & 0 & b_{23} & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ X_4 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -b_{13} & 0 \\ 0 & 1 & -b_{23} & -b_{24} \\ 0 & 0 & 1 & -b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ X_4 \end{bmatrix}$$

46 パス解析モデル

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -b_{13} & 0 \\ 0 & 1 & -b_{23} & -b_{24} \\ 0 & 0 & 1 & -b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ X_4 \end{bmatrix}$$

$$\Rightarrow V \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -b_{13} & 0 \\ 0 & 1 & -b_{23} & -b_{24} \\ 0 & 0 & 1 & -b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} V(e_1) \\ V(e_2) \\ V(e_3) \\ V(X_4) \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & -b_{13} & 0 \\ 0 & 1 & -b_{23} & -b_{24} \\ 0 & 0 & 1 & -b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \right)^T$$



47

多重指標モデル

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_{42} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ F_2 \end{bmatrix} + \begin{bmatrix} 1 \\ \lambda_{21} \\ 0 \\ 0 \\ \gamma_{21} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ d_2 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ F_2 \\ F_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{21} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{42} & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma_{21} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ F_2 \\ F_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ d_2 \\ F_1 \end{bmatrix}$$



48

多重指標モデル

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ F_2 \\ F_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -\lambda_{21} \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -\lambda_{42} & 0 \\ 0 & 0 & 0 & 0 & 1 & -\gamma_{21} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ d_2 \\ F_1 \end{bmatrix}$$

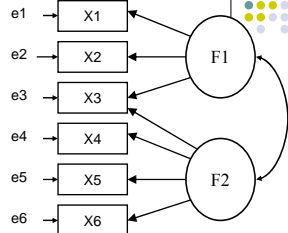
$$\Leftrightarrow \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = G \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ F_2 \\ F_1 \end{bmatrix} = G \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -\lambda_{21} \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -\lambda_{42} & 0 \\ 0 & 0 & 0 & 0 & 1 & -\gamma_{21} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ d_2 \\ F_1 \end{bmatrix}$$

$$\Rightarrow V \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = G \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -\lambda_{21} \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -\lambda_{42} & 0 \\ 0 & 0 & 0 & 0 & 1 & -\gamma_{21} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} V(e_1) \\ V(e_2) \\ V(e_3) \\ V(e_4) \\ V(d_2) \\ V(F_1) \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -\lambda_{21} \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -\lambda_{42} & 0 \\ 0 & 0 & 0 & 0 & 1 & -\gamma_{21} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \right)^T G^T$$



因子分析モデル

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ & \lambda_{21} & & & & \\ & & \lambda_{31} & & & \\ & & & \lambda_{41} & & \\ & & & & \lambda_{51} & \\ & & & & & \lambda_{61} \\ & & & & & & 1 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$



$$V \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix} + \begin{bmatrix} V(e_1) & & & & & \\ & V(e_2) & & & & \\ & & V(e_3) & & & \\ & & & V(e_4) & & \\ & & & & V(e_5) & \\ & & & & & V(e_6) \end{bmatrix}$$

$$\begin{bmatrix} V(F_1) & Cov(F_1, F_2) \\ Cov(F_2, F_1) & V(F_2) \end{bmatrix}$$



2. 適合度指標

適合度指標の意味と使い分け



適合度の評価

- モデルの適合度評価
 - SEMにおいては必須
- 適合度評価の考え方
 - データの相関行列とモデルによる相関行列の食違いの程度を評価... $S - \Sigma^{\wedge}$
 - 回帰分析では $y - y^{\wedge}$
 - 多次元量を1次元に縮約するため多種多様な方法



適合度の評価

- どの指標を使うべきか
 - 50種類以上ある
 - 一般的なコンセンサスはない
 - 開発者は自分が開発した指標を薦める
 - 回帰分析における R^2 においても百家争鳴
 - 査読者に求められたら素直に報告する
- どの指標においても、そこそこの適合が得られていることが必要



いくつかの「側面」

- 相対評価 vs 絶対評価
 - 最小モデルを導入する
 - 独立モデル, ゼロモデル
- 自由度を考慮するか
 - けちの原理 (parsimony)
 - 自由度にくらべてどの程度の適合かをみる



好ましい性質

- nに依存しないことが望ましい
- 自由度を考慮した方がよい?
 - 自由度dが小さい(パラメータが多い)モデルは適合が良いのはあたりまえ
 - しかし, 回帰分析でも R^2 が活きている
- 最小モデルを導入した方がよい?
 - 回帰分析では「一般平均のみ」の最小モデルを導入している
 - 指標が最小モデルの選択に依存する



比較

	自由度による調整		
	なし	比	差
絶対評価	カイ 2乗値	RMSEA	F0, m _k AIC
相対評価			
$\Sigma_B = O$	GFI	AGFI	
$\Sigma_B = D$	NFI IFI	NNFI RFI	CFI

指標の定義

	自由度による調整		
	なし	比	差
絶対評価	$\chi^2 = nF$	$RMSEA = \sqrt{F/d-1/n}$	$F0 = F - d/n, m_k = e^{-1/2(F-d/n)}$ (AIC = $nF - 2d$)
$\Sigma_B = O$	$GFI = \frac{B_o - F}{B_o}$	$AGFI = \frac{B_o/d_o - F/d}{B_o/d_o}$	
$\Sigma_B = D$	$NFI = \frac{B_D - F}{B_D}$	$NNFI = \frac{B_D/d_D - F/d}{B_D/d_D - 1/n}$	$CFI = \frac{(B_D - d_D/n) - (F - d/n)}{B_D - d_D/n}$
	$IFI = \frac{B_D - F}{B_D - d/n}$	$RFI = \frac{B_D/d_D - F/d}{B_D/d_D}$	

$F = \log |\hat{\Sigma}| - \log |S| + \text{tr}[\hat{\Sigma}^{-1}(S - \hat{\Sigma})], \frac{1}{2} \text{tr} \left[\left\{ (S - \hat{\Sigma}) \hat{\Sigma}^{-1} \right\}^2 \right], d = p(p+1)/2 - q$
 $B_D = \log |\text{Diag}(S)| - \log |S|, \frac{1}{2} \text{tr} \left[\left\{ (S - \hat{D}) \hat{\Sigma}^{-1} \right\}^2 \right], d_D = p(p+1)/2 - p = p(p-1)/2$
 $B_o = \frac{1}{2} \text{tr} \left[\left\{ S \hat{\Sigma}^{-1} \right\}^2 \right], d_o = p(p+1)/2$

補足:カイ2乗検定について

- 統計的検定にもとづく方法論
 - 統計的モデル評価の基本
 - H0: モデルが正しい H1: モデルが正しくない
- 専門家の意見
 - Do not rely only on the chi-square test
Bollen & Long (1993). Testing Structural Equation Models. Sage: CA (page.8)

カイ2乗検定の問題点

- 検定力がnに依存する
 - 第二種の過誤を評価していない
 - 第一種の過誤を問題とする通常の検定と逆
- 統計モデルは厳密にはデータの発生機構を表さない
- データ分布に関する基礎仮定が厳しい
 - 非正規分布の下でも正しいことがある
 - 仮定: 相関しない因子は互いに独立に分布する
- nが小さいときはカイ2乗近似が正しくなく、大きいときは殆ど確実にモデルを棄却する

3. 共分散構造分析における統計的推測

⇒ 次ページから